PROCESS FOR THE MICROBIAL PRODUCTION OF AMINO ACIDS BY BOOSTED ACTIVITY OF EXPORT CARRIERS

BACKGROUND OF THE INVENTION

The invention relates to a process for the microbial production of amino acids, to export genes, to regulator genes, to vectors,to transformed cells, to membrane proteins, and to uses.

Amino acids are of high economical interest and there are many applications for the amino acids: for example, L-lysine as well as L-threonine and L-tryptophan are needed as feed additives, L-glutamate as seasoning additive, L-isoleucine, and L-tryosine in the pharmaceutical industry, L-arginine and L-isoleucine as medicine or L-glutamate and L-phenylalanine as a starting substance for the synthesis of fine chemicals.

A preferred method for the manufacture of these different amino acids is the biotechnological manufacture by means of microorganisms; since, in this way, the biologically effective and optically active form of the respective amino acid is directly obtained and simple and inexpensive raw materials can be used. As microorganisms, for example, Carynebacterium glutamicum and its relatives ssp. flavum and ssp lactofermentum (Liebl et al; Int. J-System Bacteriol (1991) 41:255-260) as well as Escherichia coli and related bacteria can be used.

However, these bacteria produce the amino acids only in the amounts needed for their growth such that no excess amino acids are generated and are available. The reason for this is that in the cell the biosynthesis of the amino acids is controlled in various ways. As a result, different methods of increasing the formation of products by overcoming the control mechanisms are already known. In these processes, for example, amino acid analogs are utilized to render the control of the biosynthesis ineffec-

tive. A method is described, for example, wherein Coarynebacterium strains are used which are resistant to L-tyrosine and L-phenylalanine analogs (JP 19037/1976 and 39517/1978). Also methods have been described in which bacteria resistant to L-lysine and also to L-threonine analogs are used in order to overcome the control mechanisms (EP 0 205 849 B1, UK patent application GB 2 152 509 A).

Furthermore, microorganisms constructed by recombinant DNA techniques are known wherein the control of the biosynthesis has also been eliminated by cloning and expressing the genes which code for the key enzymes which cannot be feed-back inhibited any more. For example, a recombinant L-lysinee producing bacterium with plasmid-coded feedback-resistant aspartate kinase is known (EP 0381527). Also, a recombinant L-phenylalanine producing bacterium with feedback resistant prephenate dehyrodgenase has been described (JP 124375/1986; EP 0 488 424). In addition, increased amino acid yields have been obtained by overexpression of genes which do not code for feedback-sensitive enzymes of the amino acids synthesis. For example, the lysine formation is improved by increased synthesis of the dihydrodipicolinate synthase (EP 0 197 335). Also, the threonine formation is improved by increased synthesis of threonine dehydratase (EP 0 436 886 A1).

Further experiments for increasing the amino acid production aim at an improved generation of the cellular primary metabolites of the central metabolism. In this connection, it is known that the overexpression of the transketolase achieved by recombinant techniques improve the product generation of L-tryptophan, L-tyrosine or L-phenalanine (EP 0 600 463 A2). Furthermore, the reduction of the phosphenol pyruvate carboxylase activity in Corynebacterium provides for an improvement in the generation of aromatic amino acids (EP 0 331 145).

All these attempts to increase the productivity have the aim to overcome the limitation of the cytosolic synthesis of the amino

acids. However, as a further limitation basically also the export of the amino acids formed in the interior of a cell into the culture medium should be taken into consideration. As a result, it has been tried to improve this export and, consequently, the efficiency of the amino acid production. For example, the cell permeability of the Carynebacterium has been increased by biotin deficiency, detergence or penicillin treatment. However, these treatments were effective exclusively in the production of glutamate, whereas the synthesis of other amino acids could not be improved in this manner. Also, bacteria strains have been developed in which the activity of the secretion system is increased by chemical or physical mutations. In this way, for example, a Corynebacterium glutamicum strain has been obtained which has an improved secretion activity and is therefore especially suitable for the L-Lysine production. (DE 02 03 320).

Altogether, the attempts to increase the secretion of amino-acids formed within the cell have all in common that an increase efflux of amino acids on the basis of the selected non-directed and non-specific methods could be achieved only accidentally.

Solely in the German patent application No. 195 23 279.8-41, a process is described which provides for a well-defined increase of the secretion of amino acids formed internally in a cell by increasing the expression of genes coding for the import of amino acids. The understanding on which this process was based, that is, the cell utilizes import proteins for the export of amino acids as well as the fact that by nature microorganisms do not generate and release excess amino acids lets one assume that export genes or proteins specific for the amino acid transport do not exist, but that the amino acids are excreted by way of other export systems.

The export systems known so far export poisonous metal ions, toxic antibiotica and higher molecular toxins. These export systems are relatively complex in their structure. Generally, mem-

brane proteins of the cytoplasmic membrane are involved which however cause only a partial reaction of the export so that presumably additional extra cytoplasmic support proteins are needed for the transport (Dink, T. et al., A family of large molecules across the outer membranes of gram-negative bacteria., J. Bacteriol. 1994, 176: 3825-3831). Furthermore, it is known that, with the sec-dependent export system for extra-cellular proteins, at least six different protein components are essential for the export. This state-of-the-art suggests that also the systems, which are responsible for the export of amino acids, but which are not known so far comprise several protein components or respectively, several genes are responsible for the export of amino acids. A hint in this direction could be the various mutants which are defective in the lysine export as described by Vrylic et al.,(J. Bacteriol (1995) 177:4021-4027).

SUMMARY OF THE INVENTION

It has now been found surprisingly that only a single specific gene is responsible for the export of amino acids so that, in accordance with the invention, for the first time a method for the microbial manufacture of amino acids is provided wherein clearly the export gene expression and/or the export carrier activity of a microorganism producing amino acids is increased. The increased export expression or respectively, activity of the export carrier resulting from this process leads to an increased secretion rate so that the export of the respective amino acid is increased. The microorganisms so modified also accumulate an increased part of the respective amino acid in the culture medium.

For an increase in the export carrier activity especially the endogenic activity of an amino acid producing microorganism is increased. An increase of the enzyme activity can be obtained for example by an increased substrate consumption achieved by changing the catalytic center or by eliminating the effects of enzyme inhibitors. An increased enzyme activity can also be caused by an

increased enzyme synthesis for example by gene amplification or by eliminating factors which inhibit the enzyme biosynthesis. The endogene export activity is increased preferably by mutation of the endogenic export gene. Such mutations can be generated either in an uncontrolled manner in accordance with classic methods as for example by UV irradiation or by mutation causing chemicals or in a controlled manner by gene-technological methods such as deletion(s) insertion(s) and/or nucleotide exchange(s).

The export gene expression is increased by increasing the number of gene copies and/or by increasing regulatory factors which positively affect the export gene expression. For example, a strengthening of regulatory elements takes place preferably on the transcription level by increasing particularly the transcription signals. This can be accomplished for example in that, by changing the promoter sequence arranged before the structure gene, the effectiveness of the promoter is increased or by completely replacing the promoter by more effective promoters. An amplification of the transcription can also be achieved by accordingly influencing a regulator gene assigned to the export gene as will be explained further below. On the other hand, an amplification of the translation is also possible, for example, by improving the stability of the m-RNA.

To increase the number of gene copies the export gene is installed in a gene construct or, respectively, in a vector, preferably, a vector with a small number of copies. The gene construct includes regulatory gene sequences, which are specifically assigned to the export gene, preferably such sequences which reinforce the gene expression. The regulatory gene sequences comprise a nucleotide sequence which codes for the amino acid sequence given in table 1 or the allele variations thereof or respectively, a nucleotide sequence 954 to 82 according to table 2 or a DNA sequence which is effective essentially in the same manner.

Allele variations or, respectively, equally effective DNA

sequences comprise particularly functional derivatives which can be obtained by deletion(s) insertion(s) and/or substitution(s) of nucleotides of corresponding sequences, wherein however the regulator protein activity or function is retained or even increased. In this way, the effectiveness of the interaction of the regulatory protein to the DNA of the export gene to be regulated can be influenced by mutating the regulatory gene sequence such that the transcription is strengthened and, consequently, the gene expression is increased. In addition, also so-called enhancers may be assigned to the export gene as regulatory sequences whereby, via an improved correlation between RNA polymerase and DNA, also the export gene expression is increased.

For the insertion of the export gene into a gene construct, the gene is preferably isolated from a microorganism strain of the type Corynebacterium and, with the gene construct including the export gene, a microorganism strain, especially Corynebacterium, producing the respective amino acid is transformed. The isolation and transformation of the respective transport gene occurs according to the usual methods. If a transport gene is isolated and cloned from Corynebacterium then for example, the method of homologuous complementation of an export defective mutant is suitable (J.Bacteriol. (1995)177: 4021-4027). If a direct cloning of the structure gene is not possible vector sequences may first be inserted into the transport gene whereupon it is isolated by way of "plasmid rescue" in the form of inactive fragments. For the process according to the invention genes from the C. glutamicum ATCC 13032 or C. glutamicum ssp. flavum 14067 or also, C. glutamicum ssp. lacto fermentum ATCC 13869 are particularly suitable. The isolation of the genes and their in-vitro recombination with known vectors (Appl. Env. Microbial (1989)55: 684-688; Gene 102(1991)93-98) is followed by the transformation into the amino acid producing strains by electroporation (Liebl et al. (1989)FEMS Microbiol Lett. 65; 299-304) or conjugation (Schäfer et al. (1990) J. Bacte-

riol. 172:1663-1666). For the transfer, preferably vectors with low numbers of copies are used. As host cells, preferably such amino acid producers are used which are deregulated in the synthesis of the respective amino acids and/or which have an increased availability of central metabolism metabolites.

After isolation, export genes with nucleotide sequences can be obtained which code for the amino acid sequences given in table 3 or for their allele variations or, respectively, which include the nucleotide sequence of 1016 to 1725 according to table 2 or a DNA sequence which is effective essentially in the same way. Also here, allele variations or equally effective DNA sequences include particularly functional derivatives in the sense indicated above for the regulatory sequences. These export genes are preferably used in the process according to the invention.

One or several DNA sequences can be connected to the export gene with or without attached promoter or respectively, with or without associated regulator gene, so that the gene is included in a gene structure.

By cloning of export genes, plasmids or, respectively, vectors can be obtained which contain the export gene and which, as already mentioned, are suitable for the transformation of an amino acid producer. The cells obtained by transformation which are mainly transformed cells from Corynebacterium, contain the gene in reproducible form, that is, with additional copies on the chromosome wherein the gene copies are integrated at any point of the genome by homologous recombination and/or on a plasmid or respectively, vector.

A multitude of sequences is known which code for membrane proteins of unknown function. By providing in accordance with the invention export genes such as the export gene with the nucleotide sequence of nucleotide 10165 to 1725 in accordance with table 2 or respectively, the corresponding export proteins for example that with the amino acid sequence according to table 1, it is now pos-

sible to identify by sequence comparison membrane proteins, whose function is the transport of amino acids. The export gene identified in this way can subsequently be used to improve the amino acid production in accordance with the process of the invention.

The membrane proteins known from the state-of-the-art generally include 12, some also only 4 transmembrane helices. However, it has now been found surprisingly that the membrane proteins responsible or suitable for the export of amino acids include 6 transmembrane helices (see for example, the amino acid sequence of an export protein listed in the table 3, wherein the 6 transmembrane areas have been highlighted by underlining). Consequently, there is a new class of membrane proteins present which has not yet been described.

BRIEF DESCRIPTION OF DRAWINGS

Figure 1 shows the fragments in pMV6-3 and pMV8-5-24 obtained by the cloning which cause the lysine secretion and the subclone pMV2-3 made from pMV6-3, which also causes the lysine secretion and which was sequenced. B_1BamH1; Sm, SmaI; Se, Sac1; S1, Sal I,II, HindII; X, XhoI.

Figure 2 shows a comparison of the derivated amino acid sequence of LysE from C. glutamicum (above), with a gene product of so far unknown function from Escherichi coli (below), which is identified thereby as export carrier.

Fig. 3 shows increased lysine export by pMV2-3 with C. glutamicum NA8. On top, the control with low excretion and cell-internal backup of lysine up to about 150mM. Below, the high secretion caused by pMV2-3 with cell internally only small backup of about 30mM.

Figure 4 shows the increase of the lysine accumulation in C. glutamicum by lys E lys G(pMV2-3) (middle curve), and the accumulation caused by lysE(plysE) (upper curve).

Examples:

a) Cloning of an export gene and cloning of a regulator of

Corynebacterium glutamicum.

Chromosomal DNA from C. glutamicum R127 (FEMS Microbiol lett. (1989)65:299-304) was isolated as described by Scharzer et al. (Bio/Technology (1990) 9:84-87). The DNA was then split with the restriction enzyme Sau3A and separated by saccharose gradient centrifugation as described in Sambrook et al. (Molecular cloning, A laboratory manual (1989) Cold Spring Harbour Laboratory Press). The various fractions were analyzed gel electrophoretically with respect to their size and the fraction with a fragment size of about 6 - 10kb was used for the ligation with the vector pJC1. In addition, the vector pJC1 was linearized with BamHI and dephosphorylized. Five ng thereof was ligated with 20ng of the chromosomal 6-10 kb fragments. With the whole ligation preparation, the export defective mutant NA8 (J. Bacterol. (1995)177:4021-4027) was transformed by electroporation (FEMS Microbiol Lett(1989)65:299 - 304). The transformants were selected for LBHIS(PEMS Microbiol. Lett. (1989)65:299-304) with 15µg kanamycin per ml. These transformants were subjected to extensive plasmid analyses in that 200 of the altogether 4500 clones obtained were individually cultivated and their plasmid content and size was determined. On average, about half of the kanamycin-resistant clones carried a recombinant plasmid with an insert of the average size of 8kb. This provides for a probability of 0.96 for the presence of any gene of C. glutamicum in the established gene bank. The 4500 obtained transformants were all individually checked for renewed presence of lysinee secretion. For this purpose, the system described by Vrljic for the induction of the L-lysinee excretion in Corynebacterium glutamicum was utilized (J. Bacteriol (1995) 177:4021-4027). For this purpose, so-called minimal-medium-indicator plates were prepared, which contained per liter 20g $(NH_4)_2SO_4$, 5g uric acid, 1g KH_2PO_4, 1 g K_2HPO_4, 0.25g $MgSO_4x7H_2O$, 42 g morpholino propane sulfonic acid, 1ml $CaCl_2$ (1g/100ml), 750 ml dest., 1 ml Cg trace salts, 1 ml biotin (20µg/100l), pH7, 4% glucose, 1.8mg pro-

tocatechuic acid, 1 mg $FeSO_4$ x 7 H_2O, 1 mg $MnSO_4$ x H_2O, 0.1 mg $ZnSO_4$ x $7H_2O$, 0.02mg $CuSO_4$, 0.002mg $NiCl_2$ x $6H_2O$, 20 g agar-agar, as well as 10^7 cells/ml of the lysine-auxotrophene C. glutamicum mutant 49/3. The original 4500 transformants were all individually pinned, by toothpicks onto the indicator plates with, in each case, a check of the original non-excretor NA8 (J.Bacteriol (1995)177:4021-4027) and the original strain R127. At the same time, always 2 plates were inoculated of which only one contained additionally 5mM L-methionine in order to induce the lysine excretion in this way. The indicator plates were incubated at 30°C and examined after 15, 24 and 48 hours. In this way, altogether 29 clones were obtained which showed on the indicator plate provided with methionine a growth court by the indicator strain 49/3. The clones were examined individually and then again as described above, for reestablishment of the growth court. In this way, the two clones NA8 pMV8-5-24 and NA8 pMV6-3 were obtained which had again received the capability to excrete lysine.

From these clones, plasmid preparations were performed as described in Schwarzer et al. (Bio/Technology (1990)9; 84-87). By retransformation in NA8, the plasmid-connected effect of the excretion of L-lysine was confirmed. Both plasmids were subjected to a restriction analysis. Plasmid pMV8-5-24 carries an insert of 8.3 kb, and pMV6-3 one of 9.5 kb. The physical charter of the inserts is shown in Fig. 1.

b) Subcloning of an DNA fragment which reconstitutes the lysine export.

From the insert of the plasmid pMV6-3 individual subclones were prepared utilizing the restriction severing point as determined. In this way, the 3.7 kb XhoI-SalI-fragment, the 2.3 kb BamHI-fragment and the 7.2 kb BamHI fragment were ligated with the correspondingly severed and treated vector pJC1 (Mol Gen. Genet.(1990)220: 478-480). With the ligation products C. glutamicum NA8 was directly transformed, the transformants were

tested for having the lysine excretion properties and the presence of the subclone was confirmed by plasmid preparation and restriction analysis. In this way, the strain with plasmid pMV2-3 (Fig. 1) was obtained as smallest subclone. This fragment resulting in lysine export contains as insert the 2.3kb Bam fragment from pMV6-3.

c) Sequence of the lysine export gene lys E and its regulators lysG.

The nucleotide sequence of the 2.3kb BamH1 fragment was performed according to the dideoxy-chain termination method of Sanger et al. (Proc. Natl. Acad. Sci USA(1977) 74:5463-5467) and the sequencing reaction with the Auto Read Sequencing kit from Pharamcia (Uppsala, Sweden). The electrophoretic analysis occurs with the automatic laser-fluorescence DNA sequencing apparatus (A.L.F) from Pharmacia-LKB(Piscataway, NJ, USA). The nucleotide sequence obtained was analyzed by the program packet HUSAR (Release 3.0) of the German Cancer Research Center (Heidelberg). The nucleotide sequence and the result of the analysis is presented in Fig. 2. The analysis results in two fully open reading frames (ORF) on the sequenced DNA piece. ORF1 codes for a protein with a length of 236 amino acids, OFR2 codes for a protein with a length of 290 amino acids. The protein derived from ORF1 includes an accumulation of hydrophobic amino acids as they are characteristic for membrane-embedded proteins. The detailed analysis of the distribution of the hydrophobic and hydrophilic amino acids by the programs PHD.HTM (Protein Science(1995)4:521-533) is shown in table 3. It is apparent therefrom that the protein contains six hydrophobic helix areas which extend through the membrane. Consequently, this protein is the searched for exporter of the amino acid L-lysine. The corresponding gene will therefore be designated below as lysE. In table 2, it is marked accordingly. ORF2 is transcribed in a direction opposite to ORF1. The sequence analysis shows that ORF2 has a high identity with regulator genes

which are combined as a single family (Ann Rev Microbiol(1993) 597-626). Genes of this family regulate the expression processes of the various genes involved in catabolic or anabolic processes in a positive way. For this reason, ORF2 will below be designated as lysG (Govern=regulating). Because of the coordination and because lysE could be cloned (see a)) and subcloned (see b)) together with lysG, lysG is regulator of lysE and consequently also participates in the lysine export. The gene lysG and the amino acid sequence derived therefrom are also shown in table 2 and, respectively, table 1.

d) Identification of an unknown membrane protein from Escherichia coli by sequence comparison.

With the established sequences according to table 3, already existing sequence banks can be searched in order to assign the proteins derived in this way from sequenced areas a certain function. Correspondingly, the amino acid sequence of the lysine exporters consisting of C. glutamicum were compared with derivated protein sequences of all the DNA sequences deposited there utilizing the program packet HUSAR (Release 3.0) of the German Cancer Research Center (Heidelberg). A high homology of 39.3% identical amino acids and 64.9% similar amino acids was found to a single sequence of so far unknown function of E.coli.

The comparison is shown in Fig. 2. The open read frame of E. coli so far not characterized is consequently identified by way of this process as an amino export gene.

e) Increased export of intracellularly accumulated L-lysine.

The strain C. glutamicum NA8 (J. Bacteriol(1995) 177: 4021-4027 was transformed with plasmid pMV2-3 and the L-lysine excretion of the strains was compared. For this purpose, NA8 and NA8pMV2-3 in complex medium were utilized as described in Vrljic et al. (J. Bacteriol (1995)177:4021-40277) and the fermentation medium CGXII (Bacteriol (1993)175:5595-5603 were each separately inoculated. The medium additionally contained 5mM L-methionin in

order to induce the intracellular L-lysine biosynthesis. After cultivation for 24 hours at 30°C on a rotary vibrator at 140 rpm, the cell internal and external L-lysine determinations were performed. For the cell-internal determination silicon oil centrifugations were performed (Methods Enzymology LV(1979) 547-567); the determination of the amino acids occurred by high pressure liquid chromatography (J. Chromat (1983) 266:471-482). These determinations were performed at different times as indicated in Fig. 3. In accordance with the process used the retained cell internal L-lysine is excreted also by pMV2-3 to a greater degree and is accumulated. Accordingly, also the cell internally present L-lysine is greatly reduced. Consequently, the utilization of the newly discovered and described exporter represents a process for greatly improving the L-lysine production.

f) Increased accumulation of L-lysine by lysE or LysEG.

From the subclone pMV2-3 which contains the sequenced 2374bp Bam HI-fragment in pJCI (see figure 1), the lysE carrying 1173 bpPvuII fragment was ligated in pZ1 (Appl. Env. Microbiol(1989)55:684-688) according to the sequence information and in this way, the plasmid plysE was obtained. This plasmid as well as the lysE lysG carrying plasmid pMV2-3 was introduced into C. glutamicum strain d by electroporation wherein the chromosomal areas were deleted. The obtained strains C. glutamicum d pMV2-3, C. glutamicum d plysE, C. glutamicum pJC1 were, as described under e) precultivated on a complex medium, then cultivated in production minimal medium CGx11 together with 4% glucose and 5mM l-methionin and samples were taken to determine the accumulated lysine. As apparent from Fig. 4 with lysE lysG an increase of the lysine accumulation with respect to a control sample is achieved. With plysE an extraordinarily increased accumulation of from 4.8 to 13.2 mM L-lysine is achieved with this method.

LEGENDS OF THE TABLES

Table 1: The amino acid sequence of the lysine exporter regulator from Corynebacterium glutamicum with the helix-turn-helix motive typical for DNA-binding proteins.

Table 2(three pages): The nucleotide sequence of C. glutamicum coding for the lysine exporter and lysine export regulators.

Table 3: the amino acid sequence of the lysine exporter from Corynebacterium glutamicum with the identified transmembrane helices TMH1 to TMH6.

Table 1

1 MNPIQLDTLL SIIDEGSFEG ASLALSISPS AVSQRVKALE HHVGRVLVSR

Helix-Turn-Helix-Motiv

51 TQPAKATEAG EVLVQAARKM VLLQAETKAQ LSGRLAEIPL TIAINADSLS

101 TWFPPVFNEV ASWGGATLTL RLEDEAHTLS LLRRGDVLGA VTREANPVAG

151 CEVVELGTMR HLAIATPSLR DAYMVDGKLD WAAMPVLRFG PKDVLQDRDL

201 DGRVDGPVGR RRVSIVPSAE GFGEAIRRGL GWGLLPETQA APMLKAGEVI

251 LLDEIPIDTP MYWQRWRLES RSLARLTDAV VDAAIEGLRP

Table 2

```
                                                                          60
GGTAAACGACTTCCACAATGAGACGGACCGGGTTAAGGACGCCCGCTTCTTCACTTTTTG

                                                                         120
                    ⊢LysG
GGACTTGGAAAAGTCTTCATTGATTCCGGCGTTAGGGAGCTAACGACGTAGTTGCTGCCG
                    -  P  R  L  G  E  I  A  A  D  V  V  A

                                                                         180
CAGACACTCAGATCGATCTCTAGATCTAAGGTCCGCGGTAGCAACGGTTATGTAGCCACA
 D  T  L  R  A  L  S  R  S  E  L  R  W  R  Q  W  Y  M  P  T

                                                                         240
CAGTTACCCATAGAGTAGCTCCTCCTAGTGAAGAGGACGAAAATCGTACCCTCGTCGAAC
 D  I  P  I  E  D  L  L  I  V  E  G  A  K  L  M  P  A  A  Q

                                                                         300
CCAAAGCCCTTCTTCAGGGGTTGGTTCCGGAGCCGCTTAACGGAGTGGTTTTGGAAGGCG
 T  E  P  L  L  G  W  G  L  G  R  R  I  A  E  G  F  G  E  A

                                                                         360
GCTGCCCTGTTACCTATGCGCGGACGCGGGGTGTCCTGGTAGCTGCGCGGGCAGGTCCAG
 S  P  V  I  S  V  R  R  R  G  V  P  G  D  V  R  G  D  L  D

                                                                         420
TGCCAGAACTTCGTGTAGAAACCCTGGCTTCGCATTCTGCCCGTAGCGTCGGGTTAGATC
 R  D  Q  L  V  D  K  P  G  F  R  L  V  P  M  A  A  W  D  L

                                                                         480
AAAGGGTAGTTGGTACATCCGTAGGGCGTTACTCCCCCAACGTTACCGGTTCACCGCGTA
 K  G  D  V  M  Y  A  D  R  L  S  P  T  A  I  A  L  H  R  M

                                                                         540
CCAAGGTTCAAGATGATGAAGTGTAGGGCGGTGCCCTAATCGAAGTGCCCAATGGCGAGG
 T  G  L  E  V  V  E  C  G  A  V  P  N  A  E  R  T  V  A  G

                                                                         600
ATTTTGTAGAGGTGCGGCGTCGTTCCTATTACACACGCGAAGTAGAAGGTTCGCGTCGCA
 L  V  D  G  R  R  L  L  S  L  T  H  A  E  D  E  L  R  L  T

                                                                         660
CTCGCAACGAGGTGGGGTTCTTCGATGGAGCAACTTGTGCCCTCCTTTGGTACACCTATC
 L  T  A  G  G  W  S  A  V  E  N  F  V  P  P  F  W  T  S  L

                                                                         720
GCTTAGACGCAACTACCGCTACCAATTGCCCTAAAGTCGTTCCGCAGGTCTATCAACGCG
 S  D  A  N  I  A  I  T  L  P  I  E  A  L  R  G  S  L  Q  A

                                                                         780
AAATCAAAGACGAACGTCGTTGTGGTAAAAGGCGCGACGAACGTGTTCCTGAAGTGGGCG
 K  T  E  A  Q  L  L  V  M  K  R  A  A  Q  V  L  V  E  G  A

                                                                         840
AAGCCAACGAAACCGGCCAACCCACGCGCTATGGTTGTGAGCTGGGTGCACTACGAGCTC
 E  T  A  K  A  P  Q  T  R  S  V  L  V  R  G  V  H  H  E  L

                                                                         900
TCGAAATTGCGCGACTGAGTGGCGGCTCCCCCTTTACCTTTCCCGATTCCTCCGCGGAAG
 A  K  V  R  Q  S  V  A  S  P  S  I  S  L  A  L  S  A  G  E
```

Table 2 (cont.)

```
                                                                   RCGS           960
                                                                  <---LysG
CTTCGACGGAAGTAGTTACTAACTCTCGTTTCACAGGTCAACTTACCCCAAGTA-----5'
5'---TGCCTTCATCAATGATTGAGAGCAAAGTGTCCAGTTGAATGGGGTTCATGAAGCT
   F  S  G  E  D  I  I  S  L  L  T  D  L  Q  I  P  N  M

                                                             RBS          1020
ATATTAAACCATGTTAAGAACCAATCATTTTACTTAAGTACTTCCATAGGTCACGATGGT
                                                       M  V
                                                    LysE--->

          .         .         .         .         .         .         1080
GATCATGGAAATCTTCATTACAGGTCTGCTTTTGGGGGCCAGTCTTTTACTGTCCATCGG
 I  M  E  I  F  I  T  G  L  L  L  G  A  S  L  L  L  S  I  G

          .         .         .         .         .         .         1140
ACCGCAGAATGTACTGGTGATTAAACAAGGAATTAAGCGCGAAGGACTCATTGCGGTTCT
 P  Q  N  V  L  V  I  K  Q  G  I  K  R  E  G  L  I  A  V  L

          .         .         .         .         .         .         1200
TCTCGTGTGTTTAATTTCTGACGTCTTTTTGTTCATCGCCGGCACCTTGGGCGTTGATCT
 L  V  C  L  I  S  D  V  F  L  F  I  A  G  T  L  G  V  D  L

          .         .         .         .         .         .         1260
TTTGTCCAATGCCGCGCCGATCGTGCTCGATATTATGCGCTGGGGTGGCATCGCTTACCT
 L  S  N  A  A  P  I  V  L  D  I  M  R  W  G  G  I  A  Y  L

          .         .         .         .         .         .         1320
GTTATGGTTTGCCGTCATGGCAGCGAAAGACGCCATGACAAACAAGGTGGAAGCGCCACA
 L  W  F  A  V  M  A  A  K  D  A  M  T  N  K  V  E  A  P  Q

          .         .         .         .         .         .         1380
GATCATTGAAGAAACAGAACCAACCGTGCCCGATGACACGCCTTTGGGCGGTTCGGCGGT
 I  I  E  E  T  E  P  T  V  P  D  D  T  P  L  G  G  S  A  V
                           >>>>>>> >       < <<<<<<

          .         .         .         .         .         .         1440
GGCCACTGACACGCGCAACCGGGTGCGGGTGGAGGTGAGCGTCGATAAGCAGCGGGTTTG
 A  T  D  T  R  N  R  V  R  V  E  V  S  V  D  K  Q  R  V  W

          .         .         .         .         .         .         1500
GGTAAAGCCCATGTTGATGGCAATCGTGCTGACCTGGTTGAACCCGAATGCGTATTTGGA
 V  K  P  M  L  M  A  I  V  L  T  W  L  N  P  N  A  Y  L  D

          .         .         .         .         .         .         1560
CGCGTTTGTGTTTATCGGCGGCGTCGGCGCGCAATACGGCGACACCGGACGGTGGATTTT
 A  F  V  F  I  G  G  V  G  A  Q  Y  G  D  T  G  R  W  I  F

          .         .         .         .         .         .         1620
CGCCGCTGGCGCGTTCGCGGCAAGCCTGATCTGGTTCCCGCTGGTGGGTTTCGGCGCAGC
 A  A  G  A  F  A  A  S  L  I  W  F  P  L  V  G  F  G  A  A

          .         .         .         .         .         .         1680
AGCATTGTCACGCCCGCTGTCCAGCCCCAAGGTGTGGCGCTGGATCAACGTCGTCGTGGC
 A  L  S  R  P  L  S  S  P  K  V  W  R  W  I  N  V  V  V  A
```

Table 2 (cont.)

```
              .         .         .         .         .         .  1740
                    >>>>>>>       <<<<<<<    |- orf3
                                             -  N  E  R  T  K
5'     CTACTGGCGTAACCGGTAGTTTGACTACAACTACCCAATCAAAAGCGCCCAAAA
AGTTGTGATGACCGCATTGGCCATCAAACTGATGTTGATGGGTTAGTTTTCGCGGG    5'
 V  V  M  T  A  L  A  I  K  L  M  L  M  G  -
                                     LysE -|        >>>>>

              .         .         .         .         .         .  1800
CCTTAGCCACCGGAAGCGGGTTTACAACTACGGCCGCAGCACCCTTTAGAGTAGCTAGCG
 S  D  T  A  K  A  W  I  N  I  G  A  D  H  S  I  E  D  I  A
              <<<<<

              .         .         .         .         .         .  1860
GAGGTTGAGCCGCAGTCTTTTGAGGTTCAACAACTCACTTAGTTCCGACAACAGGTCGAC
 E  L  E  A  D  S  F  E  L  N  N  L  S  D  L  S  N  D  L  Q

              .         .         .         .         .         .  1920
GAGTTGACTGCTTCGTGGTTAGTTACGTGACCAGTGCCATAGGCGCGGCATGAGAGGAAC
 E  V  S  S  A  G  I  L  A  S  T  V  T  D  A  G  Y  E  G  Q

              .         .         .         .         .         .  1980
GAGCGCGTCGTGGGTACGTTCGCGGTAGACGCGTTCACTGACGGGCGCAAGGACCCGCTA
 E  R  L  V  W  A  L  A  M  Q  A  L  S  Q  G  R  E  Q  A  I

              .         .         .         .         .         .  2040
CAGTAACTCGAACGCCTGGTATAGTTATAACAAGTGCAAGTTGTACGGGAGTCTGTCCCT
 D  N  L  K  R  V  M  D  I  N  N  V  N  L  M  G  E  S  L  S

              .         .         .         .         .         .  2100
GAATGGGACCGACCGCGCCCTTGGGAGACCTTAAGGTAGCTCTATAAACAGGCACTCGTC
 K  G  Q  S  A  R  S  G  E  P  I  G  D  L  Y  K  D  T  L  L

              .         .         .         .         .         .  2160
CGGGACGCGTTCACCACTCTTTCGTTACTGCGGTTCTGGTAACAACCGTCGACTGACGTT
 G  Q  A  L  P  S  F  A  I  V  G  L  G  N  N  A  A  S  Q  L

              .         .         .         .         .         .  2220
GTTCAAGAGTGGCAGTAGCGGGCCAAGGAGGTGGGTTGCTAATTACTACCTTATCGAACC
 L  N  E  G  D  D  G  P  E  E  V  W  R  N  I  I  S  Y  S  P

              .         .         .         .         .         .  2280
GACTACTTAGTCTTCGCCCGTCGGGAGGAGGCGGTACTTGAGTCGGCGGAGGCGACACTC
 Q  H  I  L  L  P  C  G  E  E  A  M  F  E  A  A  E  A  T  L

              .         .         .         .         .         .  2340
GAGACCTGGCATCCTTCTTTATGGGTGCATTTCTCGGAAAGGTCTGCGTTGTTACAGTGC
 E  P  G  Y  S  S  I  G  V  Y  L  A  K  G  S  A  V  I  D  R

              .         .         .         .         .         .  2374
                                         <-orf3-|
GTTACGCATGTACCAAAGAAGGTTTCCTCATAGA
 L  A  Y  M  T  E  E  L  P  T  D
```

Table 3

1 MVIMEIFITG LLLGASLLLS IGPQNVLVIK QGIKREGLIA VLLVCLISDV

TMH1 **TMH2**

51 FLFIAGTLGV DLLSNAAPIV LDIMRWGGIA YLLWFAVMAA KDAMTNKVEA

TMH3

101 PQIIEETEPT VPDDTPLGGS AVATDTRNRV RVEVSVDKQR VWVKPMLMAI

151 VLTWLNPNAY LDAFVFIGGV GAQYGDTGRW IFAAGAFAAS LIWFPLVGFG

TMH4 **TMH5**

201 AAALSRPLSS PKVWRWINVV VAVVMTALAI KLMLMG

TMH6

Add a1